Available online at www.sciencedirect.com

SCIENCE DIRECT®

LINEAR ALGEBRA AND ITS APPLICATIONS

Linear Algebra and its Applications 386 (2004) 207–223

www.elsevier.com/locate/laa

Finding equilibrium probabilities of QBD processes by spectral methods when eigenvalues vanish

Winfried K. Grassmann*

Department of Computer Science, University of Saskatchewan, 57 Campus Drive, Engineering Building, Saskatoon, Canada SK S7N 5A9

Received 12 August 2003; accepted 7 December 2003

Submitted by D. Szyld

Abstract

In this paper, we discuss the use of spectral or eigenvalue methods for finding the equilibrium probabilities of quasi-birth–death processes for the case where some eigenvalues are zero. Since this leads to multiple eigenvalues at zero, a difficult problem to analyze, we suggest to eliminate such eigenvalues. To accomplish this, the dimension of the largest Jordan block must be established, and some initial equations must be eliminated. The method is demonstrated by two examples, one dealing with a tandem queue, the other one with a shorter queue problem.

Keywords: Quasi-birth–death process; Tandem queues; Shorter queue; Markov chains; Eigenvalues

1. Introduction

In this paper, we deal with continuous-time Markov chains having the following block-structured matrices

* Tel.: +1-306-966-4898; fax: +1-306-966-4884.
E-mail address: grassman@cs.usask.ca (W.K. Grassmann).

0024-3795/$ - see front matter
doi:10.1016/j.laa.2003.12.013

$$Q = \begin{bmatrix} A_{b,b} & A_{b,0} & 0 & \cdots & \\ A_{0,b} & A_{0,0} & A_0 & 0 & \ddots \\ 0 & A_2 & A_1 & A_0 & \ddots \\ 0 & \ddots & \ddots & \ddots & \end{bmatrix}. \tag{1}$$

Markov chains of this type are known as quasi-birth–death processes, or QBD processes. The Markov chain contains the levels $b, 0, 1, 2, \ldots$, and the transition matrix is partitioned according to levels. Each level contains a number of phases. Level b is the boundary level, and it contains N_b, phases. All other levels contain N phases. The block $A_{b,b}$ reflects transitions occurring from a state to level b to a state of the same level, the block $A_{b,0}$ contains the rates of going from level b to level 0, and the block $A_{0,b}$ contains the rates of going from level 0 to b. The block $A_{0,0}$ must be understood in a similar fashion. The blocks A_j contain the rates of changing the level by $1 - j$, $j = 0, 1, 2$. $A_{0,0}$ and all A_j are square matrices of dimension N.

The objective is to find the equilibrium probabilities of the process. Hence, let π_n be the vector of all equilibrium probabilities of level n, $n = b, 0, 1, \ldots$. We assume that the process is positively recurrent. As pointed out in [6], given a matrix of the form of (1), one can always eliminate π_b, the boundary probabilities, obtaining a infinitesimal generator of the form

$$Q = \begin{bmatrix} A_{0,0} & A_0 & 0 & \cdots \\ A_2 & A_1 & A_0 & \ddots \\ 0 & \ddots & \ddots & \ddots \end{bmatrix}. \tag{2}$$

Note that $A_{0,0}$ in (2) is different from $A_{0,0}$ in (1). To indicate this we replace $A_{0,0}$ in (1) by $A'_{0,0}$, and we determine the $A_{0,0}$ in (2) as

$$A_{0,0} = A'_{0,0} + A_{0,b}(-A_{b,b})^{-1}A_{b,0}. \tag{3}$$

The A_i, $i = 0, 1, 2$ do not change. As the reader may have guessed, this result can be obtained by block elimination. As shown by [11], block elimination results in a new Markov chain with all states eliminated censored, which is to say that the sample function while in these states is removed. It follows that the equilibrium probabilities corresponding to (2) are proportional to the ones corresponding to (1). Hence, without loss of generality, we can assume that the matrix in question has been brought into the form given by (2). The equilibrium equations can now be written as:

$$0 = \pi_0 A_{0,0} + \pi_1 A_2, \tag{4}$$

$$0 = \pi_n A_0 + \pi_{n+1} A_1 + \pi_{n+2} A_2, \quad n \geqslant 0. \tag{5}$$

Moreover, if e is the column vector with all its entries equal to 1, we have

$$\sum_{n=0}^{\infty} \pi_n e = 1. \tag{6}$$

According to Neuts [14], the probabilities π_n are given by

$$\pi_n = \pi_0 R^n, \tag{7}$$

where R is the minimal non-negative matrix satisfying

$$0 = A_0 + RA_1 + R^2 A_2. \tag{8}$$

It can be shown that all eigenvalues of R are inside the unit circle.

Our focus is on the method of spectral analysis, or eigenvalue method. In this method, one finds values x_i (the eigenvalues), and vectors $d^{(i)} \neq 0$ (the eigenvectors) that satisfy the following equation (see [7,9,12] and references in these papers)

$$0 = d^{(i)}(A_0 + A_1 x_i + A_2 x_i^2) = d^{(i)} A(x_i),$$

where $A(x) = A_0 + A_1 x + A_2 x^2$. One then forms the diagonal matrix $\Lambda = \mathrm{diag}(x_i)$ and the matrix $D = [d^{(i)}]^{\mathrm{T}}$. Hence, row i of D is equal to $d^{(i)}$. If all eigenvalues are distinct, then one has

$$\pi_n = c\Lambda^n D. \tag{9}$$

The vector c can be obtained from (4), (6) and (9) as follows (see also [9]). Eqs. (4) and (9) yield

$$0 = \pi_0 A_{0,0} + \pi_1 A_0 = cDA_{0,0} + c\Lambda DA_0$$

or

$$0 = c(DA_{0,0} + \Lambda DA_0). \tag{10}$$

Eqs. (6) and (9) yield

$$1 = c\,\mathrm{diag}(1/(1 - x_i))De. \tag{11}$$

Many researchers claim that eigenvalue methods lead to numerical instabilities, and that these instabilities are magnified if there are multiple eigenvalues. The eigenvalue $x = 0$ frequently has a multiplicity greater than 1. However, this paper shows how the resulting numerical difficulties can be bypassed. The mathematical difficulty of our method should not be minimized, but compared to a rigorous treatment of multiple eigenvalues and the corresponding Jordan chains (see e.g. in [4]), our derivation is relatively straightforward.

Generally, it has been our experience that eigenvalue solutions do not normally lead to numerical difficulties (see [6]). Moreover, we will show that for the methods discussed in this paper, potential problems are easily recognizable by considering the components of the vector c, we call them c_i. In fact, inaccuracies can only occur when some c_i are extremely large.

Eigenvalue methods seem to have an edge over matrix analytic methods regarding the computational complexity. For instance, numerical investigations done by Haverkort and Ost [10] and by Mitrani and Chakka [12] indicate that the solution times when using eigenvalue methods are substantially shorter than the ones needed for matrix analytic solutions. Moreover, we recently showed [5, 7] that if the matrices

A_0, A_1 and A_2 are all tridiagonal, eigenvalue methods have a lower time-complexity than matrix analytic methods. Eigenvalue methods can also help one to discover analytical solutions as indicated in [9].

The outline of this paper is as follows: we first discuss the geometric and algebraic multiplicity of the eigenvalues of R and $A(x)$, together with their Jordan blocks. These concepts are then used to deal with the initial conditions. The theory is then applied to solve two examples, one dealing with a tandem queue, the other one dealing with the a shorter queue model. Questions of numerical accuracy and computational complexity are then addressed. The last section presents some conclusions.

2. The eigenvalues of R and $A(x)$ and their multiplicity

The eigenvalues of $A(x)$ are, by definition, the roots of the polynomial $\det A(x)$. The eigenvalues of R, on the other hand, are the roots of the polynomial $\det(R - Ix)$. These definitions reflect the general practice, according to which the definition of an eigenvalue is different, depending whether one deals with a matrix polynomial or a matrix.

We now show that the eigenvalues of the matrix R defined by (8) coincide with the eigenvalues of $A(x)$ that are inside the unit circle. To prove this, we need the following result:

Theorem 1. *There is an invertible matrix Y and a stochastic matrix G such that*

$$A(x) = A_0 + A_1x + A_2x^2 = (R - Ix)Y(Gx - I). \tag{12}$$

For the proof of this theorem, see [13].

We have to distinguish between geometric and arithmetic multiplicity. The *geometric* multiplicity of an eigenvalue x_i is the number of independent eigenvectors connected with x_i. We denote this number by $m(x_i)$. Hence, if $DA(x_i) = 0$, where D is a rectangular matrix of rank r, and there is no matrix of rank exceeding r, then $m(x_i) = r$. The algebraic multiplicity of the eigenvalue x_i, $\widetilde{m}(x_i)$, is the multiplicity of the root x_i of the polynomial $\det A(x)$. The algebraic multiplicity can exceed the geometric multiplicity.

Theorem 2. *All eigenvalues of $A(x)$ inside the unit circle are eigenvalues of R, and they have the same arithmetic multiplicity. Similarly, all eigenvalues of $A(x)$ on or outside the unit circle are the reciprocals of the eigenvalues of G. Conversely, if x_i is an eigenvalue of R, then x_i is an eigenvalue of $A(x)$, and if y_i is a non-zero eigenvalue of G, $1/y_i$ is an eigenvalue of $A(x)$.*

Proof. Take the determinant of both sides of (12) to find

$$\det A(x) = \det(R - Ix)\det Y \det(Gx - I).$$

Clearly, if $x_i \neq 0$ is an eigenvalue of $Gx - I$, then $1/x_i$ must be an eigenvalue of G. Also, the eigenvalues of R are strictly inside the unit circle, and since G is a stochastic matrix, its eigenvalues must be on or inside the unit circle. If $x_i = 0$, then it clearly must be an eigenvalue of R, because $\det(Gx - I)$ is ± 1 in this case. If $x_i \neq 0$, and $|x_i| < 1$, it must be an eigenvalue of R, because $1/x_i$ is outside the unit circle and therefore cannot be an eigenvalue of G. Similarly, if x_i is outside the unit circle, it cannot be an eigenvalue of R, and must therefore be an eigenvalue of G. Hence, $R - Ix$ and $Gx - I$ cannot satisfy the same eigenvalue. It follows that the algebraic multiplicity $\widetilde{m}(x_i)$, $|x_i| < 1$, must be the same for R and $A(x)$. □

Theorem 3. *The geometric multiplicity of the eigenvalue x_i, $|x_i| < 1$ of $A(x)$ is equal to the geometric multiplicity of x_i of R.*

Proof. If x_i is an eigenvalue of $A(x)$ with a geometric multiplicity of $m(x_i)$, then there is a matrix $\overline{D}$ with rank $m(x_i)$, and no matrix with rank greater $m(x_i)$ satisfying $\overline{D}A(x_i) = 0$, and, since $\det(Gx_i - I) \neq 0$ and $\det Y \neq 0$, $\overline{D}(R - Ix_i) = 0$, and the result follows. □

0 is an eigenvalue of $A(x)$ if $0 = dA(0) = dA_0$ has a non-zero solution vector d, that is, A_0 must be singular. The geometric multiplicity $m(0)$ of the eigenvalue $x = 0$ is given by the dimension of the null-space of A_0. By Theorem 3, $m(0)$ is also the geometric multiplicity of the eigenvalue $x = 0$ of R, or the dimension of the null-space of R.

The matrix R, like every other matrix, can be factored as

$$R = D^{-1}\Lambda D, \tag{13}$$

where D is a non-singular matrix and Λ has the eigenvalues of R on its diagonal. If x_i is an eigenvalue with a geometric multiplicity of $m(x_i)$, there must be $m(x_i)$ *Jordan blocks* corresponding to x_i. A Jordan block of dimension 1 is merely a block containing the eigenvalue x_i as its only entry. A Jordan block of dimension greater than 1 has the following form

$$\begin{bmatrix} x_i & 1 & 0 & \cdots & \cdots & 0 \\ 0 & x_i & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & x_i & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & x_i \end{bmatrix}.$$

The algebraic multiplicity of x_i is equal to the sum of the dimensions of the Jordan blocks belonging to x_i, which is equal to the number of times x_i appears on the diagonal.

We will use the terms *zero-based eigenvector* to denote an eigenvector belonging to an eigenvalue of $x = 0$, and *zero-based Jordan block* to denote a Jordan block

with 0 on the diagonal. If J is a zero-based Jordan block of dimension d, then $J^n = 0$ $n \geqslant d$ as is easily verified. Hence, if κ is the size of the largest zero-based Jordan block, then all zero-based Jordan blocks in $R^n = D^{-1}\Lambda^n D$, $n \geqslant \kappa$, become zero. As a consequence, the null-space of R^n has dimension $\widetilde{m}(0)$ for $n \geqslant \kappa$.

To find κ, we first consider the Jordan-structure of R, and we then connect this to the Jordan structure of $A(x)$ by means of (12). The factor $Y(Gx - I)$ in this equation cannot be singular for $|x| < 1$. Consequently, for $|x| < 1$, the null-space of $A(x)$ is equal to the null-space of R. Furthermore, because of (13), we have

$$R - Ix = D^{-1}\Lambda D - Ix = D^{-1}(\Lambda - Ix)D.$$

If one is interested in the dimension of the null-space of R, one can ignore the factors of D and D^{-1} and concentrate on the matrix $\Lambda^{(0)}(x) = \Lambda - Ix$.

To eliminate eigenvalues of 0, we divide as many columns of $\Lambda(x)$ as possible by x in such a way that the entries remain polynomials. By inspection, we note that such a division can be done for every column corresponding the first column of a zero-based Jordan block. For Jordan blocks with a dimension greater 1, this division leads to the following result:

$$\begin{bmatrix} -1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & -x & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & -x & 1 \\ 0 & \cdots & \cdots & \cdots & 0 & -x \end{bmatrix}.$$

A zero-based Jordan block of dimension 1 will become a matrix with -1 as its only entry and it will remain that way. The other Jordan blocks are manipulated such that further divisions by x become possible. To this end, eliminate the 1 in row 1, column 2 of all blocks involved by adding the first to the second column of the block. After this operation, the first diagonal entry of each block is equal to -1, and the entry beside it is zero. At this point, further divisions by x become possible. This process can be repeated. To formulate this properly, consider the two sequences of matrix polynomials $\Lambda^{(k)}(x)$ and $\overline{\Lambda}^{(k)}(x)$. Both sequences start with $\Lambda^{(0)}(x) = \overline{\Lambda}^{(0)}(x) = \Lambda - Ix$. The matrices $\overline{\Lambda}^{(k)}(x)$ are obtained from the $\Lambda^{(k)}(x)$ by creating columns that have an x on the diagonal, and zeros elsewhere. All these columns can be divided by x, and the matrix created in this fashion is $\Lambda^{(k+1)}(x)$. The process stops as soon as all zero-based Jordan blocks have converted every x on the diagonal to a -1, and this requires exactly κ iterations. Hence, we obtain the following algorithm:

Algorithm 1

1. $k = 0$
2. $\Lambda^{(0)}(x) = R - Ix$

3. Repeat steps 3.1 through 3.4
   3.1 Convert $\Lambda^{(k)}(x)$ to $\overline{\Lambda}^{(k)}(x)$ by creating as many columns as possible with x as their only entries. Let $m^{(k)}$ the number of columns created this way
   3.2 If $m^{(k)} = 0$, exit loop
   3.3 Divide each of the $m^{(k)}$ columns of $\overline{\Lambda}^{(k)}(x)$ having x as its only entry by x to find $\Lambda^{(k+1)}(x)$
   3.4 $k = k + 1$
4. $\kappa = k$

We modify this algorithm to obtain Algorithm 2, which can be applied to find κ from $A(x)$.

Algorithm 2

1. $k = 0$
2. $A^{(0)}(x) = A(x)$
3. Repeat steps 3.1 through 3.4
   3.1 Convert $A^{(k)}(x)$ to $\bar{A}^{(k)}(x)$ by creating as many columns as possible with entries that are polynomials with no constant term. Let $m^{(k)}$ the number of columns created this way
   3.2 If $m^{(k)} = 0$, exit loop
   3.3 Divide each of the $m^{(k)}$ columns of $\bar{A}^{(k)}(x)$ with no constant term by x to find $A^{(k+1)}(x)$
   3.4 $k = k + 1$
4. $\kappa = k$

Algorithm 2 parallels Algorithm 1 it leads to the same value κ.

Step 3.1 of Algorithm 2, where polynomial entries with no constant terms are created, will now be fleshed out in more detail. The constant terms of the polynomial entries are given by the matrix $A^{(k)}(0)$, and we have to create a matrix $\bar{A}^{(k)}(0)$ that has the same null-space as $A^{(k)}(0)$, but that has $m^{(k)}$ columns that are 0. This can be done by adding and subtracting the different columns such that all constants of the columns to be divided by x are eliminated. The process essentially creates the echelon form of $A^{(k)}(0)$. As introductory textbooks on linear algebra show (see e.g. [15]), this is equivalent to post-multiplying $A^{(k)}(0)$ with a non-singular matrix, say $B^{(k)}$. The other steps of the algorithm should be obvious.

The remainder of this section gives additional information about Algorithm 2. It is more technical, and can be skipped at first reading. We note that whenever we divide $m^{(k)}$ columns by x, the degree of the polynomial $\det \bar{A}^{(k)}(x)$ is reduced by $m^{(k)}$. It follows that after Algorithm 2, the degree of $\det A(x)$ is reduced by $\widetilde{m} = \Sigma_{k=0}^{\kappa-1} m^{(k-1)}$. Note that any solution $(d^{(i)}, x_i)$ still satisfies $0 = d^{(i)} A^{(k)}(x_i)$ as long as $x_i \neq 0$. To see this, write

$$A^{(k+1)}(x) = A^{(k)}(x) B^{(k)} I^{(k)}(1/x),$$

where $I^{(k)}(1/x)$ represents the divisions by x in iteration k. For $x \neq 0$, the matrix $I^{(k)}(1/x)$ is well-defined and non-singular, and if $0 = d^{(i)}A^{(k)}(x_i)$ is a solution, $x_i \neq 0$, so is

$$0 = d^{(i)}A^{(k+1)}(x_i) = d^{(i)}A^{(k)}(x_i)B^{(k)}I^{(k)}(1/x_i).$$

Of course, $0 = d^{(i)}A(x_i) = d^{(i)}A^{(0)}(x_i)$, and by induction, $0 = d^{(i)}A^{(k)}(x_i)$ for all values of k. Moreover, since $A^{(\kappa)}(0)$ is not singular, $A^{(\kappa)}(x)$ has no longer any eigenvalue of $x = 0$. □

3. The elimination of the zero-based eigenvectors

Even in case of multiple eigenvalues, one can still use Eqs. (9), (10) and even (11) provided diag$(1/(1 - x_i)$ is replaced by $(I - \Lambda)^{-1}$. However, finding D becomes difficult for $\kappa > 1$, and we therefore use a different method. According to (9), $\pi_n = c\Lambda^n D$. For $n \geqslant \kappa$, the nth power of all zero-based Jordan blocks become 0. To reflect this, partition Λ as follows

$$\Lambda = \begin{bmatrix} \Lambda_0 & 0 \\ 0 & \Lambda_1 \end{bmatrix}.$$

Here, Λ_0 is the square matrix containing all zero-based Jordan blocks, and Λ_1 contains all other Jordan blocks. Hence,

$$\Lambda^n = \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_1^n \end{bmatrix}, \quad n \geqslant \kappa.$$

We partition c and D conformal with Λ and obtain

$$\pi_n = [\bar{c}_0 \quad \bar{c}_1] \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_1^n \end{bmatrix} \begin{bmatrix} D_0 \\ D_1 \end{bmatrix} = \bar{c}_1 \Lambda_1^n D_1, \quad n \geqslant \kappa.$$

Hence, $\bar{c}_0$ and D_0 are not needed to find $\pi_n, n \geqslant \kappa$. Moreover, if $n < \kappa$, one can find π_n directly. First, one eliminates all $\pi_n, n < \kappa$ from the equilibrium equations. Using the GTH algorithm [8], this be done in a numerically stable way. The elimination then provides equations for the backsubstitution step, and hence expressions to evaluate the $\pi_n, n < \kappa$. After eliminating $\pi_n, n < \kappa$, from the equilibrium equations, we obtain the following matrix

$$Q^{(\kappa)} = \begin{bmatrix} A_{\kappa,\kappa} & A_0 & 0 & \cdots \\ A_2 & A_1 & A_0 & \ddots \\ 0 & \ddots & \ddots & \ddots \end{bmatrix}. \tag{14}$$

From this system, we find

$$0 = \pi_\kappa A_{\kappa,\kappa} + \pi_{\kappa+1} A_2.$$

Since $\pi_n = \bar{c}_1 \Lambda_1^n D_1$, $n \geqslant \kappa$, this yields

$$0 = \bar{c}_1 \Lambda_1^{\kappa} (D_1 A_{\kappa,\kappa} + \Lambda_1 D_1 A_2).$$

We set $\hat{c} = \bar{c}_1 \Lambda^{\kappa}$, yielding

$$0 = \hat{c}(D_1 A_{\kappa,\kappa} + \Lambda_1 D_1 A_2). \tag{15}$$

In addition, the probabilities must be normed such that their sum is 1. This yields N independent equations for the $N - \widetilde{m}(0)$ variables, that is, the system is overdetermined.

However, since (10) has a solution, a solution must exist here as well, and this implies that the equations are dependent. One can thus choose a subset of these equations or, alternatively, one can use the least square method. Of course, once $\hat{c}$ is found,

$$\pi_n = \hat{c} \Lambda_1^{n-\kappa} D_1, \quad n \geqslant \kappa. \tag{16}$$

At this stage, we can summarize the procedure as follows

Algorithm 3

1. Find all non-zero eigenvalues and the corresponding eigenvectors
2. Find D_1, the matrix consisting of the non zero-based eigenvectors
3. Find κ
4. Eliminate π_n, $n < \kappa$ from the equilibrium equations. This provides expression for π_n, $n < \kappa$ in terms of π_{n+1}, and it also provides $A_{\kappa,\kappa}$
5. Solve (15) for $\hat{c}$, up to a factor f. This factor must later be found by using the condition that the sum of all probabilities equals 1
6. Use (16) to determine π_{κ} up to a factor f
7. Find π_n, $n < \kappa$, up to a factor f, by using the equations obtained in step 4 of this procedure
8. Find f by using $1 = \sum_{n=0}^{\kappa-1} \pi_n + \hat{c}(I - \Lambda)^{-1} D_1$.

This algorithm provides π_n, $n \geqslant 0$, without using any zero-based eigenvector. Hence, the problem of the multiplicity of the eigenvector $x = 0$ is bypassed.

We now show what causes the probabilities π_n, $n < \kappa$ to be different from the π_n, $n \geqslant \kappa$ given in (16). To this end, we write

$$A^{(k)}(x) = A_0^{(k)} + A_1^{(k)} x + A_2^{(k)} x^2.$$

$A^{(k)}(x)$ corresponds to an infinite matrix of the form

$$\begin{bmatrix} \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & A_2^{(k)} & A_1^{(k)} & A_0^{(k)} & 0 & \cdots \\ \vdots & 0 & A_2^{(k)} & A_1^{(k)} & A_0^{(k)} & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \ddots \end{bmatrix}.$$

To find $\bar{A}^{(k)}(x)$, columns of this matrix are added or subtracted. This addition or subtraction is done for all levels, but we never add columns from different levels. To find $A^{(k+1)}(x)$ from $\bar{A}^{(k)}(x)$, some columns are divided by x. A division of a column by x is equivalent to moving the column up one level as the reader may verify. Instead of moving up one level, one can also take the corresponding column from the previous level, and insure in this fashion that we use only the existing columns. Hence, as we go from $A^{(k)}(x)$ to $A^{(k+1)}(x)$, we borrow, so to say, from the previous level.

Consider now Q as given in (2). As in Algorithm 2, κ is the first matrix $A^{(k)}(x)$ non-singular. Since level 0 is different, we are not allowed to borrow from level 0. $A^{(\kappa)}$ at level n borrows from $n-\kappa$ and this implies that n must be greater κ in order to avoid borrowing from level 0. On the other hand, the operations of finding $A^{\kappa}(x)$ do not change the solution π_n: we only re-ordered and added equations. It follows that for level $n > \kappa$, we have

$$0 = \pi_{n-1}A_0^{(\kappa)} + \pi_n A_1^{(\kappa)} + \pi_{n+1}A_2^{(\kappa)}. \tag{17}$$

This can now be used for an alternative proof of (16). Since $A^{(\kappa)}(x)$ has all the eigenvalues and eigenvectors of $A(x)$ except the ones connected with $x = 0$, $A^{(\kappa)}(x)$ shares Λ_1 and D_1 with $A(x)$. It follows that $\pi_n = \hat{c}\Lambda_1^{n-\kappa}D_1$ for high enough n, and the only thing missing is the value of the first n satisfying this equation. Since $A_0^{(\kappa)}$ is non-singular, π_{n-1} is uniquely determined by (17) given π_n and π_{n+1}, that is, if $n > \kappa$, we have $\pi_{n-1} = \hat{c}\Lambda_1^{n-1-\kappa}D_1$ as long as π_n and π_{n+1} satisfy a similar formula. Hence, the first time π_{n-1} could be different is when $n - 1 < \kappa$, in accordance with (16).

4. Example 1: A tandem queue

We now apply our theory to solve two examples, in this section a tandem queue model, and in the next section a shorter queue model. The tandem queue considered here has two stations, both with a single exponential server. Arrivals are Poisson, and they always join the line of the first server. This line is restricted to $N-1$, and once the line is full, customers are lost. Customers having received service by the first server move on to a second line that has an unlimited capacity to wait for service by the second server. After having received service by the second server, they leave the system. In this model, the second line has to be considered the level, and the first one as the phase. The rates and the description of the different events affecting this tandem queue is given in Table 1. This table contains all possible events, together with the possible effects on level and phase. It also indicates which matrix A_i is affected by the event. For instance, an arrival increases line 1 (the phase) by 1, leaving line 2 unchanged. The arrival rate is λ, and this event can only take place if the length of the first line is less than $N-1$. Arrivals contribute to the matrix A_1.

Table 1
Tandem queue

Event	Phase	Level	Rate	Condition	A_i
Arrival	+1		λ	Phase $< N-1$	A_1
From 1 to 2	-1	+1	μ_1	Phase > 0	A_0
Departure		-1	μ_2	Level > 0	A_2

The row labeled "From 1 to 2", describes the event that a customer leaves line 1 to join line 2, decreasing thereby line 1, and increasing line 2. The rate of this event is μ_1, and it contributes to A_0. In fact, $A_0 = \mu_1 S_{-1}$, where S_{-1} is the matrix with all subdiagonal entries equal to 1, and all other entries equal to zero. The event "Departure" decreases the level by 1, has a rate μ_2 and contributes to A_2.

If $s_0 = \lambda + \mu_2$, $s = \lambda + \mu_1 + \mu_2$ and $s_{N-1} = \mu_1 + \mu_2$, $A(x)$ becomes

$$A(x) = \begin{bmatrix} -s_0x + \mu_2x^2 & \lambda x & 0 & \cdots & \cdots & \cdots & 0 \\ \mu_1 & -sx + \mu_2x^2 & \lambda x & \ddots & \ddots & \ddots & \vdots \\ 0 & \mu_1 & -sx + \mu_2x^2 & \ddots & \lambda x & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & -sx + \mu_1x^2 & \lambda x & 0 \\ \vdots & \ddots & \ddots & \ddots & \mu_1 & -sx + \mu_2x^2 & \lambda x \\ 0 & \cdots & \cdots & \cdots & 0 & \mu_1 & -s_{N-1}x + \mu_2x^2 \end{bmatrix}.$$

We now apply Algorithm 2 to find a lower bound for κ. The columns of $A(x)$ will be called column 0, column $1, \ldots,$ column $N-1$, and the rows are named in a similar way. Clearly, $A_0 = \mu_1 S_{-1}$ has rank $N-1$, and $m(0)$ is therefore 1. Hence, only one column of A_0 can be made equal to 0, and in fact, the last column (column $N-1$) of A_0 is already 0. Hence, we divide the last column of $A(x)$ by x. After the division, there are two non-zero constants in column $N-1$: λ in row $N-2$ and $-s_{n-1}$ in row $N-1$. These two constant terms can be eliminated by adding appropriate amounts of column $N-2$ and $N-3$ to column $N-1$. After this is done, two linear terms are entered into rows $N-3$ and $N-4$ of column $N-1$. We can again divide by x, and after that, these two linear terms become constants, but they, too, can be eliminated if N is large enough. In this fashion, one can continue. Each of these eliminations adds two linear terms, and after a division by x, these linear terms become constants. The process ends as soon as a constant term appears in row 0 of column $N-1$, because no elimination of such a term is possible. The number of eliminations is therefore the largest integer less than $(N-1)/2$, or $\lfloor (N-1)/2 \rfloor$. Hence, κ must be at least $\lfloor (N-1)/2 \rfloor + 1 = \lceil N/2 \rceil$. Theoretically, it could be higher because some of the constants in column $N-1$ could

have turned out to be zero, making thus additional divisions possible. This, however is extremely unlikely, and, as we will show later, it is impossible for the model considered.

In the process of finding κ, we divided $\lceil N/2 \rceil$ times by x, and this accounts for $\lceil N/2 \rceil$ eigenvalues, which leaves $N - \lceil N/2 \rceil = \lfloor N/2 \rfloor$ eigenvalues still to be determined. We will show later that all these eigenvalues are in the interval (0,1). To determine these eigenvalues, we expand $0 = dA(x)$ and obtain

$$\begin{aligned}
0 &= d_0(-s_0 x + \mu_2 x^2) + d_1\mu_1, \\
0 &= d_{i-1}\lambda x + d_i(-sx + \mu_2 x^2) + d_{i+1}\mu_1, \quad i = 1, 2, \ldots, N-2, \\
0 &= d_{N-2}\lambda x + d_{N-1}(-s_{N-1}x + \mu_2 x^2).
\end{aligned}$$

It is convenient to introduce a new variable d_N, which is zero by definition, because this allows us to write the last of the above equations as

$$0 = d_{N-2}\lambda x + d_{N-1}(-s_{N-1}x + \mu_2 x^2) + d_N\mu_1.$$

Note that for $x \neq 0$, $d_0 = 0$ implies $d_i = 0$, $i > 0$, that is, if $d_0 = 0$, the entire vector d is zero and no longer qualifies as an eigenvector. Since $d_0 \neq 0$, and since eigenvectors can only be determined up to a constant, we can set $d_0 = 1$. The equations above then lead to

$$d_1 = \frac{x}{\mu_1}(s_0 - \mu_2 x), \tag{18}$$

$$d_{i+1} = \frac{x}{\mu_1}((s - \mu_2 x)d_i - \lambda d_{i-1}), \quad i = 1, 2, \ldots, N-2, \tag{19}$$

$$d_N = \frac{x}{\mu_1}((s_{N-1} - \mu_2 x)d_{N-1} - \lambda d_{N-2}). \tag{20}$$

For any given value x, we can find the d_i recursively from (18) to (20). Of course, x is not known, and to account for this, we make the $d_i = d_i(x)$ functions of x. The problem is then to find the zeros of $d_N(x)$. The important point here is that for $x > 0$, the sequence $d_i(x)$ is a Sturm sequence. A Sturm sequence is a sequence of polynomials starting with $d_0 > 0$, where $d_{i-1}d_{i+1} < 0$ if $d_i(x) = 0$. If $n(x)$ is the number of times $d_i(x)$ changes sign as i increases, then one can prove (see e.g. [16]) that there are at least $|n(a) - n(b)|$ zeros in the interval (a, b). It is shown in [5] that $n(1-) = 0$. Furthermore, it is not difficult to show that as $x \to 0$, $d_0 = 1$, $d_i > 0$, and $d_{i+1} = -\frac{x}{\mu_1}\lambda d_{i-1}$. This means that d_{i+1} has the opposite sign of d_{i-1}, that is, there must be $\lfloor N/2 \rfloor$ sign changes, and $n(0+) = \lfloor N/2 \rfloor$. It follows that there are at least $\lfloor N/2 \rfloor$ eigenvalues in (0,1). We conclude that $\lceil N/2 \rceil$ eigenvalues are zero, and the remaining $\lfloor N/2 \rfloor$ eigenvalues are in the interval (0,1). Note that κ is a property of the tandem queue model of Table 1, and it was found without using any numerical calculation. Hence, as stated earlier, numerical stability is not an issue in this model when finding κ.

5. Example 2: A shorter queue model

In our second example, there are two exponential servers, each one having its own line. Arrivals are Poisson, and they always join the shorter line. If the two lines differ by more than $N-1$, customers switch lines. The arrival rate is λ, and each server has a service rate of μ. For formulating this model, we use the longer line as the level, and the difference between lines as the phase. Table 2 gives the details of the model. Note that A_0 has a single entry, that is, the rank of A_0 is 1, and $m(0)$, the dimension of the null-space of $A_0 = A(0)$ is therefore $N-1$. It follows that there is a single non-vanishing eigenvalue. We will show that this eigenvalue is $x_0 = \left(\frac{\lambda}{2\mu}\right)^2$. We do that by showing that $0 = dA(x_0)$ has a non-trivial solution. Similar results were obtained by Adan [1] and independently by Zhao [17] (see also [18]), but no eigenvalues were used, at least not explicitly.

We leave it to the reader to derive $A(x)$. By expanding $0 = dA(x)$ and dividing the resulting equations by x, one finds:

$$0 = -(\lambda + 2\mu)d_0 + (\lambda + \mu x)d_1, \tag{21}$$

$$0 = (\lambda/x + 2\mu)d_0 - (\lambda + 2\mu)d_1 + (\lambda + \mu x)d_2, \tag{22}$$

$$0 = \mu d_{i-1} - (\lambda + 2\mu)d_i + (\lambda + \mu x)d_{i+1}, \quad i = 2, 3, \ldots N-3, \tag{23}$$

$$0 = \mu d_{N-3} - (\lambda + 2\mu)d_{N-2} + (\lambda + 2\mu x)d_{n-1}, \tag{24}$$

$$0 = \mu d_{N-2} - (\lambda + 2\mu)d_{N-1}. \tag{25}$$

We replace all instances of x in (21)–(25) by $x_0 = \left(\frac{\lambda}{2\mu}\right)^2$ to prove that x_0 is an eigenvalue. This yields

$$0 = -(\lambda + 2\mu)d_0 + \lambda\left(1 + \frac{\lambda}{4\mu}\right)d_1, \tag{26}$$

$$0 = 2\mu\left(1 + \frac{2\mu}{\lambda}\right)d_0 - (\lambda + 2\mu)d_1 + \lambda\left(1 + \frac{\lambda}{4\mu}\right)d_2, \tag{27}$$

$$0 = \mu d_{i-1} - (\lambda + 2\mu)d_i + \lambda\left(1 + \frac{\lambda}{4\mu}\right)d_{i+1}, \quad i = 2, 3, \ldots, N-3, \tag{28}$$

Table 2
The shorter queue model

Event	Level	Phase	Rate	Condition	A_i
Arrival	+1	+1	λ	Phase $= 0$	A_0
		-1	λ	Phase > 0	A_1
Dept., lines equal		+1	2μ	Phase $= 0$	A_1
Dept. form long line	-1	-1	μ	$0 <$ Phase $< N-1$	A_2
Dept. with possible switch	-1	-1	2μ	Phase $= N-1$	A_2
Dept. from short line		+1	μ	$0 <$ Phase $< N-1$	A_1

$$0 = \mu d_{N-3} - (\lambda + 2\mu) d_{N-2} + \lambda \left(1 + \frac{\lambda}{2\mu}\right) d_{N-1}, \tag{29}$$

$$0 = \mu d_{N-2} - (\lambda + 2\mu) d_{N-1}. \tag{30}$$

Obviously, (28) is a difference equation, and we can therefore use the trial solution $d_i = y^{i-1}$, where y must satisfy

$$0 = \mu - (\lambda + 2\mu) y + \lambda \left(1 + \frac{\lambda}{4\mu}\right) y^2.$$

One solution of this equation is:

$$y = \frac{1}{2 + \lambda/(2\mu)}.$$

We try this solution and set

$$d_i = y^{i-1} = \frac{1}{(2 + \lambda(2\mu))^{i-1}}, \quad i = 1, 2, \ldots, N-2. \tag{31}$$

d_0 can now be found from (26) and d_{N-1} from (30):

$$d_0 = \lambda \left(1 + \frac{\lambda}{4\mu}\right) \Big/ (\lambda + 2\mu), \quad d_{N-1} = \frac{\mu}{\lambda + 2\mu} \left(\frac{1}{2 + \lambda/(2\mu)}\right)^{N-3}. \tag{32}$$

It is now a simple matter to verify that the solution given by (31) and (32) satisfies Eqs. (27) and (29), that is, the value of y used is the correct one. A similar result was found by Adan [1].

The approach based on difference equations discussed can be used for other problems as well, even when the eigenvalues are not known, as shown in [7] or [9]. We could even have used it to solve example 1. This would have reduced the computational complexity, but it would have increased the mathematical effort considerably.

Consider now the case where there are more than two lines in parallel. The level is given by the longest line, and the phase is some kind of enumeration of the differences between the lines. Even in this case, the only way the longer line can increase is when all lines are of equal length, that is, A_0 has only one entry, and there is therefore only one eigenvalue that is not zero. In [18], Zhao even shows, though not by using our concepts, that the number of non-zero eigenvalues remains 1 even if the servers are non-homogeneous, and even if arrivals are not Poisson.

Care must be taken to make sure that the level is chosen such that A_0 has a few entries as possible. For instance, the reader may verify that if the shorter line is used as the level, $A_0 = \lambda S_{-1}$, and the geometric multiplicity is only 1, but the algebraic multiplicity is $N - 1$!

6. Numerical considerations

In this section, we compare our methods with matrix analytic methods in terms of accuracy and computational complexity. We restrict our attention to the case where

A_0, A_1 and A_2, are tridiagonal matrices, because we have extensive experience with this case, and also because eigenvalue methods seem to have an advantage in this area. Also, both of our examples fall into this category.

For the purpose of finding the computational complexity, one must distinguish the part of the solution that requires iterations, and the part that only needs to be done once.

It is well known that the iterations in any generally applicable matrix analytic method involve matrix multiplications, which right away leads to a computational complexity of $O(N^3)$ per iteration. There does not seem to be any way to reduce the computational complexity of the matrix analytic methods when solving tridiagonal models. In contrast, the effort to find an eigenvalue in the tridiagonal case is $O(N)$ because we find the $d_i(x)$ recursively by solving equations like (18)–(20). Even if we have no vanishing eigenvalues, this leads to only N applications of a method having $O(N)$ operations, and this is equivalent to a single algorithm with a complexity of $O(N^2)$, which is significantly less than $O(N^3)$ needed when using matrix analytic methods. If some eigenvalues vanish, fewer than N eigenvalues must be found, and the computational effort is consequently reduced.

Let us now consider the complexity of the non-iterative part. For the matrix analytic methods, this involves the solution of $0 = \pi_0(A_{00} + RA_2)$ which can be done in $O(N^3)$ flops. In the eigenvalue approach without eigenvalues at zero, one must solve (10), and the time to do this is also $O(N^3)$ flops. Hence, there is no big difference here. If, however, $\kappa > 0$, then one has to eliminate κ levels, each containing N phases, and the complexity to do this is $O(\kappa N^3)$. If κ depends on N, say $\kappa = \lceil N/2 \rceil$, then we are looking at a complexity of $O(N^4)$. One concludes that for low values of κ, eigenvalues methods are advantageous, but this may change as κ increases.

Next, we have to address the issue of the precision of the results. We note that for $n \geqslant \kappa$, we have

$$\pi_n = \sum_{i=1}^{N-\tilde{m}(0)} \hat{c}_i d^{(i)} x_i^{n-\kappa}. \tag{33}$$

The question is now how well these values satisfy the equilibrium equations. We claim that this is a fair comparison because when determining R, one also stops as soon as (8) is satisfied with a high enough accuracy. We know that each solution $(d^{(i)}, x_i)$ satisfies equations similar to the equations from (18) to (20) at a high accuracy, and this means that the equilibrium equation are satisfied with the same accuracy. Even if the equations $0 = d^{(i)} A(x_i)$ hold with high accuracy, it may happen that the $d^{(i)}$ are not accurate. However, since we are not interested in the eigenvectors per se, this does not matter. If the individual solutions $(d^{(i)}, x_i)$ satisfy the equilibrium equations, and if the $\hat{c}_i$ are not large, then (33) must satisfy the equilibrium equations well. For further discussion of this issue, together with numerical result, see [6]. Hence, possible problem solutions can be recognized by values of $\hat{c}_i$ that are large compared to the entries of π_n, and we have never met such a case. Hence, contrary of what one might expect, numerical stability does not seem an issue here.

7. Conclusion

Eigenvalues of zero occur whenever A_0 is singular, and A_0 is singular whenever after an increase of the level, some phases cannot be reached. In addition to the examples discussed here, there are many others. For instance, if the level is the number in any queue that is increased through Erlang-k arrivals, then after an arrival, only one phase is possible, which reduces the number of non-vanishing eigenvalues by a factor of k. For further examples of problems with vanishing eigenvalues, see [2,3].

A criticism leveled against the use of eigenvalue methods is that they may involve eigenvalues of multiplicity greater than 1, which is considered harmful. The eigenvalue $x = 0$ often has a multiplicity greater one, but in this particular case, the mathematical and numerical problems one would expect can be bypassed: One merely eliminates all eigenvalues $x = 0$, and obviously, things that are eliminated no longer exist and can therefore do no harm. To do the elimination, one has to find κ, the size of the largest Jordan block having $x = 0$ as an eigenvalue. Finding κ is typically done for entire classes of models, and it does not involve specific rates, and hence its determination is not a numerical problem, but a mathematical one. In conclusion, this paper has presented tools one can use to defuse the problems caused by multiple eigenvalues at zero.

Acknowledgements

The authors thanks the referees for their valuable suggestions. Contract/grant sponsor: Research supported in part by NSERC Disvovery Grant; contract/grant number: 8112.

References

[1] I.J.B.F. Adan, The compensation approach for queueing problems, Ph.D. Thesis, Technische Universiteit Eindhoven, 1991.
[2] A.S. Alfa, Discrete time queues and matrix-analytic methods, TOP 10 (2) (2002) 147–210.
[3] A.S. Alfa, Combined elapsed time and matrix-analytic method for the discrete time $GI/G/1$ and the $GI^X/G/1$ systems, Queueing Syst. Theory Appl. 45 (2003) 5–25.
[4] I. Gohberg, P. Lancaster, L. Rodman, Matrix Polynomials, Academic Press, New York, 1982.
[5] W.K. Grassmann, Real eigenvalues of certain tridiagonal matrix polynomials, with queueing applications, J. Linear Algebra Appl. 342 (2002) 93–106.
[6] W.K. Grassmann, The use of eigenvalues for finding equilibrium probabilities of certain Markovian two-dimensional queueing problems, INFORMS J. Comput. 15 (2003) 412–421.
[7] W.K. Grassmann, S. Drekic, An analytical solution for a tandem queue with blocking, Queueing Syst. 36 (2000) 221–235.
[8] W.K. Grassmann, M. Taksar, D.P. Heyman, Regenerative analysis and steady state distributions for Markov chains, Oper. Res. 33 (1993) 1107–1117.
[9] W.K. Grassmann, J. Tavakoli, A tandem queue with a movable server: an eigenvalue approach, SIAM J. Matrix Anal. Appl. (2002) 465–474.

[10] B.R. Haverkort, A. Ost, Steady-state analysis of infinite stochastic Petri nets: a comparison between the spectral expansion and the matrix-geometric method, in: Proceedings of the 7th International Workshop on Petri Nets and Performance Models, Saint Malo, France, IEEE Computer Society Press, 1997, pp. 36–45.
[11] J.G. Kemeny, J.L. Snell, A.W. Knapp, Denumerable Markov Chains, Van Nostrand, Princeton, NJ, 1966.
[12] I. Mitrani, R. Chakka, Spectral expansion solution for a class of Markov models: application and comparison with the matrix-geometric method, Perform. Evaluat. 23 (1995) 241–260.
[13] V. Naoumov, Matrix-multiplicative approach to quasi-birth-and-death processes analysis, in: Matrix-Analytic Methods in Stochastic Models, Marcel Dekker, New York, 1996, pp. 87–106.
[14] M.F. Neuts, Matrix-Geometric Solutions in Stochastic Models, Johns Hopkins University Press, Baltimore, 1981.
[15] G. Strang, Linear Algebra and Its Applications, second ed., Academic Press, New York, 1980.
[16] H.W. Turnbull, Theory of Equations, fifth ed., Oliver and Boyed, Edingurgh, 1952.
[17] Y. Zhao, Shortest queue models, Ph.D. Thesis, University of Saskatchewan, 1990.
[18] Y. Zhao, W.K. Grassmann, Queueing analysis of the jockeying model, Oper. Res. 43 (1995) 520–529.